DOCUMENT RESUME

ED 393 900 TM 024 894

AUTHOR Barrett, Thomas J.
TITLE An Evaluation Approach to Program Accountability: A Melding of Qualitative and Quantitative Traditions.
PUB DATE Dec 95
NOTE 8p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Achievement; *Accountability; Elementary Secondary Education; *Evaluation Methods; Formative Evaluation; Measurement Techniques; Performance Based Assessment; Program Evaluation; *Qualitative Research; Summative Evaluation; *Test Use
IDENTIFIERS Authentic Assessment; Stakeholders

ABSTRACT

The usefulness of authentic assessments as formative tools for teachers is well documented and supported. However, when educational institutions use these assessments as summative tools for evaluating programs and meeting accountability needs, several problems emerge. One set of problems relates to the difficulty of fitting authentic assessments into conventional measurement frameworks. Other problems pertain to the tendency for the tasks used for accountability to preempt other measures used by teachers to inform instruction. It is proposed that by using review teams composed of the major educational stakeholders to address accountability requirements, while at the same time assessing individual student achievement with formative measures, can meet both of these measurement needs while minimizing the impact of technical problems associated with performance assessment. (Contains six references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 393 900

TM

An Evaluation Approach to Program Accountability: A Melding of Qualitative and Quantitative Traditions

Thomas J. Barrett, Ph.D.
December, 1995

The usefulness of authentic assessments as formative tools for teachers is well documented and supported. However, when educational institutions use these assessments as summative tools for evaluating programs and meeting accountability needs, several problems emerge. One set of problems relates to the difficulty of fitting authentic assessments into conventional measurement frameworks. Other problems pertain to the tendency for the tasks used for accountability to preempt other measures used by teachers to inform instruction. It is proposed that by using review teams comprised of the major educational stake-holders to address accountability requirements, while at the same time assessing individual student achievement with formative measures that remain within the classroom, we can meet both of these measurement needs while minimizing the impact of technical problems associated with performance assessments.

Introduction

The usefulness of informal performance assessments by teachers in the classroom is seldom questioned. However, as school districts implement standardized performance assessments to either augment or replace multiple-choice tests, the problems inherent in using these measures for accountability become more apparent (Linn et. al. 1991; Mehrens, 1991; Barrett,1992; Shavelson, et. al. 1992)

The problems typically encountered fall into two main categories. First, when these tasks become part of the formal assessment system, the data summarization and reporting often lead to troublesome technical issues. Second, as Moss (1994) points out, there is a tendency for the formally administered performance assessments to take precedence over any informal measures used by teachers. As a result, pressure builds to narrow instruction to focus on the specific skills measured for accountability in much the same way that standardized multiple choice tests have tended to narrow instruction.

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
THOMAS J. BARRETT
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)"

BEST COPY AVAILABLE

TM024894

Technical Problems

As the testing and measurement community has pointed out, technical problems abound when we attempt to fit authentic assessment into traditional measurement models. These problems relate to generalizability (reliability) of results for individuals based on a limited number of tasks or different raters; ambiguity in interpreting comparisons between groups taking different tasks within a given time frame or across time frames; ambiguity concerning rater drift during holistic scoring both across short periods of time and across longer periods of time; limitations on content/process coverage or cognitive complexity of the tasks limitations on the amount of time that schools can devote to standardized externally imposed assessments; costs associated with training and implementation of centralized scoring; difficulties in scaling and equating of tasks--the list goes on and on.

In order to solve these problems, the resources needed to develop, research, and implement performance assessments can be prohibitive. Indeed, some of the hurdles that need to be overcome to allow complex performance tasks to meet the technical requirements generally associated with multiple-choice tests may simply be beyond our reach. While it may still be advisable to conduct some of these standardized performance assessments, the problems in interpreting results strongly suggests that they not be the only component of an accountability program.

Problems of Focus

In addition to the technical difficulties in using performance assessments within a conventional measurement paradigm, there is also the disturbing tendency for the assessments used for accountability to crowd out any informal assessments that teachers may otherwise use in their classrooms.

As school boards and funding agencies focus their attention on the results of generally narrow tasks comprising the "testing program" and as those district administrators charged with program oversight look to school principals for educational improvement, the kinds of performances that become the primary focus of instruction tend to be those found within the district's accountability program (Resnick and Resnick, 1992; Moss, 1994). When this happens, then the

important formative role of informal authentic assessments (e.g. portfolios, teacher observations and judgment, etc.) may also be compromised.

The Need for a Paradigm Shift

It would seem at this point that it is necessary to go beyond attempts to fit the square peg of authentic assessment into the round hole represented by conventional measurement paradigms. Instead of focusing solely on external assessment tasks for accountability, what is needed is a shift toward more judgmental approaches to assessment as described in the educational evaluation literature. The philosophical underpinning of authentic assessment seems more consistent with a constructivist/interpretive evaluation approach than with a positivist/empirical analytic approach. As suggested by evaluation theorists (Guba, et. al.), some of the qualitative research traditions are more in alignment with authentic assessment than are the quantitative traditions. However, moving exclusively to a qualitative approach leads to its own set of problems.

For instance, one might attempt to glean information from student working portfolios that could be summarized and reported for accountability purposes. However, to insure a reasonable "audit trail" the narratives describing the portfolio contents may themselves become arduous (Moss, 1994). Such attempts could also lead to an imposed standardization of portfolios which could detract from their usefulness in the classroom.

In addition, while reviews could be attempted by outside judges, it is much more consistent with the ideals of authentic assessment that such reviews be done by the person most familiar with the context of the work, i.e. the teacher. However, this can lead back to some of the problems mentioned earlier. For example, what do we do when some teachers are diligently using the portfolio process while others are implementing portfolios in a more perfunctory manner and have little in-depth knowledge of their students' work? Although there have been attempts to moderate teacher ratings with other external ratings by experts, this process gives the impression of rigor without the reality and falls back into the category of using positivist rooted methods with constructivist approaches to assessment.

Other informal measures could include teacher developed and scored performance assessments and observation instruments. But once

again, unless well trained in the safeguards used by expert qualitative researchers in their careful (and time consuming) analysis of information, the conclusions to be drawn from such data may be quite tentative.

Thus, I would argue that the exclusive use of informal, judgment-oriented assessments as the focus for accountability takes us down another problematic path and they too should not stand alone.

A Melding of Both

Instead, I believe that the best model for assessment includes both standardized, formal assessments and informal, teacher-judgment based assessments. But rather than simply aggregating and reporting separate summaries of these assessment results for accountability purposes, the information should represent pieces of evidence in a much more in-depth evaluation effort that emphasizes interpretation and context.

The suggested vehicle for accomplishing this lies in the use of review teams that periodically scrutinize many aspects of a school's program. Models already available such as high school accreditation reviews and earlier versions of Program Quality Reviews (PQR's) are consistent with the ideals of authentic assessment within qualitative and quantitative frameworks while at the same time avoiding the conflict between formative and summative assessments. By utilizing interviews, observations, reviews of diverse samples of student work, and standardized assessment summaries, a broad array of information can be condensed into a rich description of the program-including student performance-while minimizing the focus on any single assessment component.

Who to Involve?

PQR reviews are already conducted on a four year cycle which may actually be frequently enough given that most program innovations generally require a couple of years before showing measurable results anyway. By including on the team representatives of the major educational stake-holders (i.e. teachers, principal, curriculum director, assistant superintendent, board members, parents, business leaders) there would be less dependence on isolated tables of aggregated numbers to reflect educational accomplishments.

Instead, a comprehensive picture of school programs would be available to all while minimizing the potential for misrepresentation based on any of the specific components of the review. Instead of placing all the focus on the de-contextualized data emanating from a few limited assessments (i.e. NRT's, standardized performance assessments, working portfolios) these measures would be considered in context and would represent only part of the evidence being considered by the reviewers and would therefore tend not to dominate the focus of program improvement efforts.

Another attractive feature of using PQR type reviews is that since information can come from a variety of sources, it is not necessary for all grade levels and all content areas to utilize the same formal (or informal) assessments. Language arts and math might be formally assessed at one or two grades while science and history/social studies could be assessed at other grades in much the same way that it is done in some State assessment programs. In fact, by using this model, State assessments at certain designated grades might be all the formal, standardized assessment that is required. Information at other grade levels could be gathered more unobtrusively from other sources (e.g. portfolios, teacher judgment ratings, observations by review team members, interviews, etc.). One of the big threats to formal performance assessment programs is that the burden perceived by schools can make them fall of their own weight. By using a review team model, these assessments are not the sole accountability indicator and, therefore, can be limited in scope.

Summary

There would seem to be several advantages of this proposed approach to assessment. First, some of the technical problems associated with formal performance assessment programs would be less of a threat to the validity of the accountability system because they would represent only part of the system. Also, because these programs would only be part and parcel of a much broader assessment, they could be expected to coexist more harmoniously with informal, teacher-developed assessments that are used to evaluate individual student accomplishments. Secondly, the positive elements of both qualitative and quantitative assessments could be reflected in the results of the review. Finally, models of such a review process already exist and are already being implemented. All that may be additionally required is that review teams be more

broadly representative of the stake-holders in education and that the results be deemed to actually constitute the accountability program.

Currently, school districts conduct PQR's and construct comprehensive and rich assessments that provide a much more complete picture than the results of any one assessment. Yet, when asked how a school is performing, administrators tend to relegate this information to second class status and pull out a listing of norm referenced test results or results derived from a limited number of performance assessments tasks. This is tantamount to your doctor doing a thorough diagnostic evaluation, including an in-depth medical history and a diverse series of sophisticated tests, and then describing your condition by simply referring to a temperature reading!

By making use of the best information that is available to us through a comprehensive evaluation review process, any negative impact caused by the short-comings of specific assessment components will be minimized. It is a process that is defensible and because it is already being done, it avoids burdening schools with additional layers of formal assessment at a time when it is imperative that the primary focus of teachers be on the delivery of effective instruction.

Barrett, T.J. (1992, April). Technical Issues in Aggregating and Reporting Performance Assessment Results. Paper presented at the annual meeting of the American Educational Research Association, San Francisco).

Linn, R.L., Baker, E.L., Dunbar, S.B. (1991, November). Complex, Performance-Based Assessment: Expectations and Validity Criteria. Educational Researcher, V. 20, No. 8, P. 15-21.

Mehrens, W.A. (1991, April). Using Performance Assessments for Accountability Purposes: Some Problems. Paper presented at the annual meeting of the American Educational Research Association, Chicago).

Moss. P.A. (1994, March). Can there be validity without reliability? Educational Researcher, V. 23, No. 2, p 5-12.

Resnick, L.B., & Resnick, D. (1992). Assessing the thinking curriculum: New tools for educational reform. In B. Gifford and M.C. O'Connor (Eds.), Cognitive approaches to assessment. Boston: Kluwer-Nijhoff.

Shavelson, R.J., Baxter, G.P., Pine, J. (1992, May). Performance Assessments: Political Rhetoric and Measurement Reality. Educational Researcher, May, 1992, p. 22-27.